

#2 00 040  
BOC

日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

1c658 U.S. PTO  
09/846122

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
る事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
in this Office.

願 年 月 日  
Date of Application:

2000年 5月 2日

願 番 号  
Application Number:

特願2000-133973

願 人  
Applicant(s):

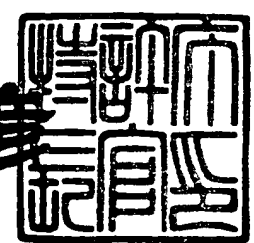
インターナショナル・ビジネス・マシーンズ・コーポレーシ  
ョン

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 6月29日

特許庁長官  
Commissioner,  
Patent Office

近 藤 隆 彦



【書類名】 特許願

【整理番号】 JP9000040

【提出日】 平成12年 5月 2日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/320

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 大和事業所内

【氏名】 根本 和郎

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【復代理人】

【識別番号】 100112520

【弁理士】

【氏名又は名称】 林 茂則

【電話番号】 046-277-0540

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【選任した復代理人】

【識別番号】 100110607

【弁理士】

【氏名又は名称】 間山 進也

【選任した復代理人】

【識別番号】 100098121

【弁理士】

【氏名又は名称】 間山 世津子

【手数料の表示】

【予納台帳番号】 091156

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識システム、音声認識方法およびコンピュータ可読な記録媒体

【特許請求の範囲】

【請求項 1】 認識単語とそれを発音表記する音素列との対応が記録された対応情報を有し、入力された音声から生成される音素列と前記対応情報内の音素列との対比によって、前記音声から一以上の認識単語を認識する音声認識システムであって、

同一人物の一定期間内の会話において、既に行われた認識処理で認識された音素列に対する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも 1 つの音素列の発声予想確率を低下させることを特徴とする音声認識システム。

【請求項 2】 同一の前記認識単語を発音表記する異なる音素列について、予め測定された発声予想確率に適合する数の前記音素列とそれに対応する前記認識単語とを前記対応情報に含める請求項 1 記載の音声認識システム。

【請求項 3】 前記一定期間は、連続した会話内の期間または当日内の複数の会話を含む期間である請求項 1 または 2 記載の音声認識システム。

【請求項 4】 認識された一以上の認識単語のうちの少なくとも一部を話者に照会して前記一以上の認識単語の認識に誤りがあるか否かを判断する手段と、  
前記誤りが認められたときには、前記一以上の認識単語をそれが誤認識されやすい一以上の認識単語に置換する手段と、

をさらに備えた請求項 1 ～ 3 の何れか一項に記載の音声認識システム。

【請求項 5】 認識された前記一以上の認識単語の数が前記認識システムにおいて予め登録された数に一致しなかったときには、前記認識された一以上の単語のうち長母音一音節を有する音素に対応する認識単語を、前記長母音に対応する短母音音素に対応する認識単語の繰り返しに置換する手段、または、前記認識された一以上の単語のうち短母音一音節を有する音素に対応する認識単語の繰り返しを、前記短母音に対応する長母音音素に対応する認識単語に置換する手段をさらに備えた請求項 1 ～ 4 の何れか一項に記載の音声認識システム。

【請求項 6】 同一人物の一定期間内の会話において、

第 1 音声を入力し、前記第 1 音声から音素列を生成するステップと、

認識単語とそれを発音表記する音素列との対応が記録された対応情報を検索し、前記第 1 音声で生成された音素列と前記対応情報内の音素列との対比から一以上の認識単語を生成するステップと、

認識された音素列に対応する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも 1 つの音素列の発声予想確率を低下させるステップと、

第 2 音声を入力し、前記第 2 音声から音素列を生成するステップと、

前記音素列の発声予想確率を低下させた対応情報を検索し、前記第 2 音声で生成された音素列と前記対応情報内の音素列との対比から一以上の認識単語を生成するステップと、

を含む音声認識方法。

【請求項 7】 同一の前記認識単語を発音表記する異なる音素列について、予め測定された発声予想確率に適合する数の前記音素列とそれに対応する前記認識単語とを前記対応情報に含める請求項 6 記載の音声認識方法。

【請求項 8】 前記一定期間は、連続した会話内の期間または当日内の複数の会話を含む期間である請求項 6 または 7 記載の音声認識方法。

【請求項 9】 認識された一以上の認識単語のうち少なくとも一部を話者に照会して前記一以上の認識単語の認識に誤りがあるか否かを判断するステップと、

前記誤りが認められたときには、前記一以上の認識単語をそれが誤認識されやすい一以上の認識単語に置換するステップと、

をさらに含む請求項 6 ～ 8 の何れか一項に記載の音声認識方法。

【請求項 10】 認識された前記一以上の認識単語の数が前記認識システムにおいて予め登録された数に一致しなかったときには、前記認識された一以上の認識単語のうち長母音一音節を有する音素に対応する認識単語を、前記長母音に対応する短母音音素に対応する認識単語の繰返しに置換するステップ、または、前記認識された一以上の単語のうち短母音一音節を有する音素に対応する認識

単語の繰り返しを、前記短母音に対応する長母音音素に対応する認識単語に置換するステップ、をさらに含む請求項 6～9 の何れか一項に記載の音声認識方法。

【請求項 1 1】 音声を認識するプログラムコードが記録されたコンピュータ可読な記録媒体であって、前記プログラムコードには、

音声を入力し、前記音声から音素列を生成するプログラムコードと、

認識単語とそれを発音表記する音素列との対応が記録された対応情報を検索し、前記音声で生成された音素列と前記対応情報内の音素列との対比から一以上の認識単語を生成するプログラムコードと、

認識された音素列に対応する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも 1 つの音素列の発声予想確率を低下させるプログラムコードと、

を含むコンピュータ可読な記録媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、音声認識システムまたは音声認識方法に関する。特に、複数の読みがある文字、単語等の認識精度の向上に適用して有効なものである。

【0 0 0 2】

【従来の技術】

たとえば、「ViaVoice（商標）」等の音声認識システムにおいては、マイクrophon、ヘッドセット等の入力デバイスにより入力された音声を音素に分解し、この解析された音素とデータベースに蓄積された音素とを対比して入力音声に対応する文字、単語等を抽出して音声認識を行っている。ここで、データベースに蓄積されている音素は個別独立的に記録されるのではなく、一定の規則に則った文法に関連付けて記録される。すなわち、たとえば図 9（a）に示すように、4 桁の数字を認識する場合には、<digits>として 4 桁の<num1>を定義し、<num1>には 0～9 のアラビア数字が入り得ることを規定する。このような文法定義の下に、図 9（b）に示すように、音素表記表を定義する。つまり、「0」には「ree」（れえ）、「ree:」（れー）、「rei」（れい）、「zero」（ぜろ）の 4 つの

音素が対応し、「1」には「ichi」（いち）の1つの音素を対応させる。その他同様に数字「2」には3音素が、「3」には1音素が、「4」には4音素が対応し、「5」以下の各数字も同様である。図9（a）の文法に図9（b）の音素表記を適用した例を図9（c）に示す。図9（c）の文法および音素表記が実用的なベースフォームとして用いられる。

## 【0003】

<digits>に対応した入力音声たとえば「ゼロいちに一さん」であれば、この音声を音素列に分けて「ゼロ」（zero）、「いち」（ichi）、「に一」（nii:）、「さん」（sa\_n）が得られ、各音素列に対応する数字「0」、「1」、「2」、「3」が音素対応表から得られる。各数字は文法定義に当てはめられ、認識結果の<digits>として「0123」の4文字が得られる。

## 【0004】

なお、「ViaVoice（商標）」等の音声認識システムでは、エンロールメントと称される処理で入力音声の個人的相違を検出し、その個人に適合した音響特性を学習して認識精度を向上する手法が適用されている。

## 【0005】

## 【発明が解決しようとする課題】

ところが、たとえば日本語数字の読み上げを考えた場合、数字音声認識の認識精度は必ずしも高いとはいえない。その原因として、以下のような事項が考えられる。

## 【0006】

まず、「いち」「に」「さん」等、日本語数字は概して単語長が短く、音の冗長性が少ないことが挙げられる。短い音の音素間にはその相違が少なく、認識における音素の取り違えが発生しやすい。日本語数字では「に」「し」「ご」「く」等の一音節のものもあり、音の冗長性の少なさは特に顕著である。

## 【0007】

次に、ひとつの数字に複数の読みが与えられていることが挙げられる。たとえば、「0」という数字には「ゼロ」「れい」「まる」等の読みが与えられ、「4」という数字には「し」「よん」等、「7」という数字には「なな」「しち」等

、「9」という数字には「きゅう」「く」という読みが与えられる。複数の読みが与えられると認識すべき音素の候補が多くなり、誤認識を誘発する確率が高くなる。

## 【0008】

また、異なる数字との間に類似する音素の多い点が挙げられる。たとえば「しち」（7）という音素と「いち」（1）という音素と「はち」（8）という音素は互いに類似しており、「し」（4）という音素と「しち」（7）という音素が類似する。また、「に」（2）、「し」（4）という音素間も類似し、「さん」（3）、「よん」（4）間も類似する。これら互いに類似する音素間では、その音の類似性により音素の識別が難しくなり、音の取り違いによる誤認識を生ずる確率が高くなる。特に電話等を介した音声識別では、入力音の低周波成分が少なくなり、低周波成分での識別が必要となるイ段の音の聞き分けが困難になる状況が生じ、この問題が顕著になる。

## 【0009】

さらに、必ずしも長母音が付くとは限らない長母音付き一音節の発音では、繰り返しが識別し難くなる問題がある。たとえば「に」「にー」「にーにー」あるいは「ご」「ごー」「ごーごー」等である。特に通常は「ごー」と発音される「5」の文字が、「753」の場合には「しちごさん」と発音されたり、「358」の場合には「さんごっば」と発音されたりして「ごー」が「ご」あるいは「ごっ」と短母音化することがあり、かえって問題を複雑にしている。

## 【0010】

ところで、数字の音声認識は、たとえば電話による取引等のビジネスの場面を考えた場合、会員番号や商品選択等の入力に多用されることが想定され、ビジネスへの展開を考慮した場合には、特に数字の音声認識精度の向上が強く求められている。

## 【0011】

なお、前記エンロールメントでは、音響特性を個人に適合させて認識精度を向上することがある程度可能であるが、入力される音素自体が類似する場合や音素に冗長性がない前記のような場合には認識精度の更なる向上には自ずと限界があ



る。

【 0 0 1 2 】

本発明の目的は、音声認識における認識精度をさらに向上することにある。特に複数の読みが与えられてる文字、単語等の認識精度を向上することにある。

【 0 0 1 3 】

また、本発明の他の目的は、冗長性の少ない音たとえば日本語における数字や、異なる文字あるいは単語に類似の音を与えられている文字等の認識精度を向上することにある。

【 0 0 1 4 】

本発明のさらに他の目的は、長母音付き一音節文字が短母音化されて繰り返される場合の認識精度を向上することにある。

【 0 0 1 5 】

【課題を解決するための手段】

本願発明は、同一人物は一回の同一会話内で一定して同じ読みを維持するという本発明者らによる知見を活用するものである。つまり、たとえば「7」を「しち」と発音した人物は、その会話内で一貫して「しち」と発音し続ける傾向がある。この傾向を活用して、会話における1回目の応答においてその人物が使わない読みに対応する音素列を認識候補からはずす、またはその認識確率を下げ、2回目以降の認識を実行するものである。

【 0 0 1 6 】

すなわち、本願発明の音声認識システムは、認識単語とそれを発音表記する音素列との対応が記録された対応情報を有し、入力された音声から生成される音素列と前記対応情報内の音素列との対比によって、前記音声から一以上の認識単語を認識する音声認識システムであって、同一人物の一定期間内の会話において、既に行われた認識処理で認識された音素列に対する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも1つの音素列の発声予想確率を低下させるものである。また、本願発明の音声認識方法は、同一人物の一定期間内の会話において、第1音声を入力し、前記第1音声から音素列を生成するステップと、認識単語とそれを

発音表記する音素列との対応が記録された対応情報を検索し、前記第 1 音声で生成された音素列と前記対応情報内の音素列との対比から一以上の認識単語を生成するステップと、認識された音素列に対応する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも 1 つの音素列の発声予想確率を低下させるステップと、第 2 音声を入力し、前記第 2 音声から音素列を生成するステップと、前記音素列の発声予想確率を低下させた対応情報を検索し、前記第 2 音声で生成された音素列と前記対応情報内の音素列との対比から一以上の認識単語を生成するステップとを含む。これにより 2 回目以降の誤認識の確率を低下して認識精度を向上できる。

## 【 0 0 1 7 】

なお、前記認識単語には字、詞または単語を含めることができ、前記対応情報には、前記認識単語が所定の規則で配列される文法情報を含めることができる。また、前記認識単語は、日本語または英語の数字、数詞または数を表す単語とすることができる。このような数字の認識において本願発明の効果は特に著しい。

## 【 0 0 1 8 】

なお、認識された音素列に対応する認識単語に複数の音素列が対応付けられて記録されている場合、前記複数の音素列のうち前記認識された音素列と異なる少なくとも 1 つの音素列の発声予想確率を低下させる態様については、前記認識された音素列と異なる音素列を前記対応情報から除外し、発声予想確率をゼロにする態様も含まれる。

## 【 0 0 1 9 】

また、同一の前記認識単語を発音表記する異なる音素列について、予め測定された発声予想確率に適合する数の前記音素列とそれに対応する前記認識単語とを前記対応情報に含めることができる。これによりさらに認識精度を高めることができる。

## 【 0 0 2 0 】

また、前記一定期間は、連続した会話内の期間または当日内の複数の会話を含む期間とすることができる。発音者の発音傾向は 1 日単位等の短時間特に同一会話内では維持される傾向が強い。しかし、逆に時間が経過した後には発音者の発

音傾向が変化することもある。このような場合本願発明では長時間の同一発音者の発音傾向を維持することなく、前記一定期間が経過した後には認識に用いる前記情報を初期状態に戻すことができる。このような操作が可能なのは、本願発明がたとえば1会話内のその場限りの学習という手法を採用する効果に基づくものである。

#### 【0021】

なお、前記一定期間内の会話が連続した会話でない場合には、パスワード、会員番号、発呼側電話番号、音声を解析して話者を特定する手法、またはこれらの組み合わせで、話者を特定する手法を用いることができる。

#### 【0022】

また、認識された一以上の認識単語のうちの少なくとも一部を話者に照会して前記一以上の認識単語の認識に誤りがあるか否かを判断し、前記誤りが認められたときには、前記一以上の認識単語をそれが誤認識されやすい一以上の認識単語に置換することができる。このような手段または方法をとることによりたとえばチェックディジット付きの長い数字の連続発声の認識精度を向上できる。

#### 【0023】

また、認識された前記一以上の認識単語の数が前記認識システムにおいて予め登録された数に一致しなかったときには、前記認識された一以上の認識単語のうち長母音一音節を有する音素に対応する認識単語を、前記長母音に対応する短母音音素に対応する認識単語の繰り返しに置換する、または、前記認識された一以上の単語のうち短母音一音節を有する音素に対応する認識単語の繰り返しを、前記短母音に対応する長母音音素に対応する認識単語に置換することができる。このような手段または方法を備えることにより長母音一音節で与えられる文字等が短母音化されて繰り返される場合、あるいは短母音の繰り返しで与えられる文字等が長母音一音節の文字等と誤認される場合の認識精度を向上できる。

#### 【0024】

#### 【発明の実施の形態】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。ただし、本発明は多くの異なる態様で実施することが可能であり、本実施の形態の記載内容に限

定して解釈すべきではない。なお、実施の形態の全体を通して同じ要素には同じ番号を付するものとする。

#### 【 0 0 2 5 】

以下の実施の形態では、主に方法またはシステムについて説明するが、当業者であれば明らかなとおり、本発明は方法、システムその他、コンピュータで使用可能なプログラムコードが記録された媒体としても実施できる。したがって、本発明は、ハードウェアとしての実施形態、ソフトウェアとしての実施形態またはソフトウェアとハードウェアとの組合せの実施形態をとることができる。プログラムコードが記録された媒体としては、ハードディスク、CD-ROM、光記憶装置または磁気記憶装置を含む任意のコンピュータ可読媒体を例示できる。

#### 【 0 0 2 6 】

本実施の形態で利用できるコンピュータシステムには、中央演算処理装置（CPU）、主記憶装置（メインメモリ：RAM(Random Access Memory)）、不揮発性記憶装置（ROM(Read Only Memory)）等を有し、これらがバスで相互に接続される。バスには、その他コプロセッサ、画像アクセラレータ、キャッシュメモリ、入出力制御装置（I/O）等が接続される。バスには、適当なインターフェイスを介して外部記憶装置、データ入力デバイス、表示デバイス、通信制御装置等が接続されてもよい。その他、一般的にコンピュータシステムに備えられるハードウェア資源を備えることが可能なことは言うまでもない。外部記憶装置は代表的にはハードディスク装置が例示できるが、これに限られず、光磁気記憶装置、光記憶装置、フラッシュメモリ等半導体記憶装置も含まれる。なお、データの読み出しのみに利用できるCD-ROM等の読み出し専用記憶装置もデータあるいはプログラムの読み出しにのみ適用する場合には外部記憶装置に含まれる。データ入力デバイスには、キーボード等の入力装置、マウス等ポインティングデバイスを備えることができる。データ入力デバイスには音声入力装置も含む。表示装置としては、CRT、液晶表示装置、プラズマ表示装置が例示できる。本実施の形態のコンピュータシステムには、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等各種のコンピュータが含まれる。

#### 【 0 0 2 7 】

各コンピュータシステムで利用されるプログラムは、他のコンピュータシステムに記録されていても良い。つまり、コンピュータシステムで利用する一部のプログラムをリモートコンピュータで分散的に処理または実行できる。なお、他のコンピュータシステムに記録されたプログラムをアドレスで参照する場合には、DNS、URL等を用ることができる。

## 【 0 0 2 8 】

## (実施の形態 1)

図 1 は、本発明の一実施の形態である音声認識システムの概要を示したブロック図である。本実施の形態の音声認識システムは、ユーザ 1 の音声が入力される音声認識エンジン 2 と、音声認識エンジン 2 に適用される文法データ（グラマー） 3 と、文法データ 3 に適用される音素表記表 4 とを有する。

## 【 0 0 2 9 】

ユーザ 1 からの音声信号は、たとえばマイクロフォン、ヘッドセット等の入力装置で電気信号に変換され、さらに A/D（アナログデジタル）変換等が行われて、デジタルデータで表された波形データに変換される。この波形データは、音声認識エンジン 2 で音素に分解され、文法データ 3 と比較されて音声認識が行われる。

## 【 0 0 3 0 】

音声認識エンジン 2 では、入力された音素に最も適合する文法データ 3 が選択される。文法データ 3 には音素表記表 4 が適用されており、入力される音声がある可能性のある文法、あるいは発声される可能性がある音素を含む形態で多数の文法データ 3（音素表記表 4 に基づいた音素が配置されている）が用意され、この文法データ 3 のデータベースを参照して入力音素に適合する文法データ 3 が選択される。

## 【 0 0 3 1 】

本実施の形態で用いる文法は図 9（a）に示すものと同等のものである。しかし、本実施の形態で用いる音素表記表は図 9（b）のものとは相違する。

## 【 0 0 3 2 】

図 2 は、本実施の形態で用いる音素表記表の一例を示す表図である。従来、た

例えば「0」の場合、文字「0」と4つの音素（「ree」「ree:」「rei」「zero」）を対応させていたのみであるが、本実施の形態では、文字「0」の読みとして「レイ」と「ゼロ」の2種類を考え、各音素に読み（「0」の場合「レイ」と「ゼロ」）の情報を付加している。なお、「0」にはほかに「マル」という読みも考えられるが、ここでは2種類の読みがあるとして例示している。

#### 【0033】

また、「4」については「シ」と「ヨン」の読みを、「7」については「シチ」と「ナナ」の読みを、「9」については「ク」と「キュウ」の読みを考える。本実施の形態では、「0」「4」「7」「9」について上記のような複数の読みを考えているがあくまでも例示であり、その他の読みが考えられるときには音素表記表に加えても良いことは勿論である。また、上記以外の数字に複数の読みが考えられるときにはその複数の読みを加えても良いことは勿論である。さらに、本実施の形態では数字を例示するが、数字に限らず、その他漢字、英字等を適用できるし、その読みに複数の候補があるときには、前記同様にそれらを含めて音素表記表を作成しても良い。

#### 【0034】

このような文法と音素表記を用いて音声認識する場合の手順を以下に説明する。図3は、本実施の形態の音声認識方法の一例を示したフローチャートである。ここでは、たとえば電話を介して取引データを音声により入力する場合を考える。

#### 【0035】

まず、本実施の形態の音声認識システムにベースとなるベース文法データを導入する（ステップ10）。導入されるベース文法データは、図2に示すように、読みの情報が付加されている以外は図9（c）と同様である。

#### 【0036】

次に、システム側の音声発声手段から「電話番号をどうぞ」というメッセージをユーザ側に流す（ステップ11）。仮に、ユーザ側の発声が「シチヨンゼロ」とすると、この音声入力に応答してベース文法データにより認識を行う（

ステップ12)。ここで、システムが「740」を認識すれば、システム側から「740で正しいですか」というメッセージを発する（ステップ13）。ユーザが「はい」と応答すれば認識結果は「740」で確定する。仮に、認識結果が誤りであれば、ステップ11に戻って再度音声入力を行う。

## 【0037】

この1回目の音声認識の結果、「740」が確定すれば、ユーザは、「7」については「シチ」と読み、「4」については「ヨン」と読み、「0」については「ゼロ」と読むことがわかる。ここで、このユーザについては、少なくとも同一会話内あるいは短時間（たとえば1日内）の会話では、同一の読みをすると仮定することが可能になる。これは、同一人物は、同一会話内で同じ読み方を維持する傾向があるという本発明者らの知見に基づく。この傾向を利用し、このユーザが「7」について「ナナ」と発音する確率は少ないと考えることができる。同様に、「4」については「シ」と、「0」について「レイ」と発音する確率は小さいと考えることができる。よって、次ステップでは、減縮した文法データを挿入する（ステップ14）。

## 【0038】

図4は、減縮した文法データを示すリスト図面である。この文法データはBNF表記による文法に相当する。すなわち、文字「7」については、「ナナ」に対応する音素「nana」を削除し、「シチ」に対応する音素「hichi」「shichi」に限定する（20）。同様に、「4」については、「シ」に対応する音素「shi」「shii」「shii:」を削除し、「ヨン」に対応する音素「yo\_n」に限定する（21）。「0」については、「レイ」に対応する音素「ree」「ree:」「rei」を削除し、「ゼロ」に対応する音素「zero」に限定する（22）。このように、発声される確率の少ない音素を削除し、この削除した音素が適用された文法データを用いて認識を行えば、より認識率を向上することができる。

## 【0039】

その後、この減縮した文法データを適用して次ステップに進み、2回目の音声入力を促すメッセージを発し（ステップ15）、音声入力に応答して認識を行う

(ステップ16)。音声入力「ゼロサンのイチシチキュウ」とすれば、本来識別が困難な「イチ」と「シチ」のような発声でも、「シチ」が「7」に限定されているので、ベース文法データを用いる場合よりも認識の精度を向上することができる。

#### 【0040】

さらに、2回目の認識でこのユーザは「9」を「キュウ」と発声してるので、音素「k u」を削除してさらに減縮した文法データを適用できる(ステップ18)。図5はさらに減縮した文法データを示すリスト図である。「9」について音素「k u」を削除している(23)。

#### 【0041】

なお、音素対応表に関連付けた文法データは、1回目の認識の前に各読みの組み合わせの数だけ予め作成しておき、2回目の認識以降で不要な文法データを削除する方法をとることができる。

#### 【0042】

本実施の形態によれば、同一人物による発声では、数字等の読み方が一貫しており、同一会話内では同じ読みが維持されるという傾向を利用し、2回目以降の認識では、読まれる可能性の低い音素を削除して認識を行える。これにより、2回目以降の認識率を向上することができる。なお、本発明は、発声者に固有な音響特性を学習するようなものではなく、その場限りの学習に止まり、他の発声者の場合や、同じ発声者でも日が異なるような場合では、再度ベース文法データを用いて最初のステップから認識が開始される。このため、発声者ごとのデータベースを作成する必要がなく、どのような発声者にも同じ構成のシステムを適用できる。このため、発声者ごとにカスタマイズする必要がなく、簡単に本発明を実施適用できる。

#### 【0043】

##### (実施の形態2)

クレジットカードのように入力される数字の特定の桁にチェックディジットが含まれるような場合には、このチェックディジットを用いて、推定される誤りを入れ替えて誤り訂正を行い、訂正後の認識結果で次ステップに進めることができる。



る。図 6 は実施の形態 2 の認識方法の一例を示したフローチャートである。

#### 【 0 0 4 4 】

まず、ベース文法データを導入し（ステップ 3 0）、カード番号の入力を促すメッセージ（ステップ 3 1）の後、ベース文法データを適用した認識を行う（ステップ 3 2）。仮に入力音声が「仔ニサソゴ`ロクツハチセ`ロ仔ニサソゴ`ロ」であったとき、「1234567890123456」と認識しない誤認識の確率は 1 回目の認識ゆえ低くはない。通常クレジットカードの場合は、チェックディジットは下 2 桁の「5 6」なので、このチェックディジットが合っているかどうかの照合を行う（ステップ 3 3）。照合の結果が合わないときには誤認識の可能性のある数字を交換して（ステップ 3 4）、再度照合を行う（ステップ 3 3）。この結果、照合にパスすれば次ステップに進め、誤り訂正後の認識結果（数字）で確認を行う（ステップ 3 5）。

#### 【 0 0 4 5 】

誤認識の可能性のある数字は、たとえば 1「イチ」と 7「シチ」と 8「ハチ」（「7」が「シチ」と発声されているときあるいは「7」が認識されていないとき）、1「イチ」と 8「ハチ」（「7」が「ナナ」と発声されているとき）、4「シ」と 7「シチ」（「4」が「シ」と発声され、「7」が「シチ」と発声されているとき）、あるいは 6「ロク」と 9「ク」（「9」が「ク」と発声されているとき）、が挙げられる。これらを相互に置換することによりステップ 3 4 の交換が行える。

#### 【 0 0 4 6 】

最後に、1 回目の認識結果と前記誤り訂正処理で確定した認識結果から、不要な音素を削除し、減縮した文法データを作成し、2 回目以降の認識のためにこれを導入する（ステップ 3 6）。

#### 【 0 0 4 7 】

本実施の形態の音声認識方法によれば、チェックディジットを用いて誤り訂正を行い、ユーザの入力回数を少なくしてユーザに対する利便性を向上し、トータルの処理時間も短くできる。

#### 【 0 0 4 8 】

## (実施の形態3)

文法で5桁の数字が要求されているにもかかわらず、4桁の数字しか認識できない場合には、長母音一音節の音声（たとえば2「n i i :」）が短母音化して2「n i」が繰り返されていた可能性がある。この場合の誤り訂正の方法が図7に示されている。図7（a）は、本実施の形態の音声認識方法の一例を示したフローチャートであり、図7（b）は適用される文法データである。図7（b）に示すように、ここでは5桁の数字が文法として要求されている（40）。

## 【0049】

図7（a）に示すように、まず、ベース文法データを導入し（ステップ41）、番号の入力を促すメッセージ（ステップ42）の後、ベース文法データを適用した認識を行う（ステップ43）。ここで、音声入力として「イチニニサンシ」が入力されたとすると、「1234」と誤認識する可能性がある。ステップ44で5文字認識できたかどうかを判断し（ステップ44）、4文字しか認識できていないときには、長母音1音節の文字「2」が誤認識されたと推定する。よって、「2」を「22」に置換し（ステップ45）、確認ステップ（ステップ46）に進む。5桁の文字が正常に認識されていたときにはそのままステップ46に進む。その後、1回目の認識結果と前記誤り訂正処理で確定した認識結果から、不要な音素を削除し、減縮下文法データを作成し、2回目以降の認識のためにこれを導入する（ステップ47）。

## 【0050】

本実施の形態の音声認識方法によれば、長母音一音節の音に誤認識されやすい短母音の繰り返しを誤り訂正することができる。これにより、ユーザの入力回数を少なくしてユーザに対する利便性を向上し、トータルの処理時間も短くできる。なお、上記実施の形態では、短母音の繰り返しが長母音の一文字に誤認された例を示したが、長母音の一文字が短母音の繰り返しの誤認されて、本来の文字数より多く誤認識された場合についても、上記と同様に誤りの訂正ができることは勿論である。

## 【0051】

以上、本発明者によってなされた発明を発明の実施の形態に基づき具体的に説

明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能である。

## 【 0 0 5 2 】

たとえば、図 8 に示すように、文法に非一様な確率分布を適用できる。すなわち、「0」についていえば、「ゼロ」と発音される確率が「レイ」と発音される確率より経験的に大きいことが判明している。このようなとき、図 8 に示すように<0>に「z e r o」の音素を 3 つ割り当て、<r e i>を 1 つ割り当てる。この場合、「セロ」と発声される確率が 7 5 %、「レイ」と発声される確率が 2 5 %と仮定している。なお、「レイ」の発声には「r e e」「r e e :」「r e i」の音素が均等に割り当てられている。このような非一様確率分布を適用することにより、より認識精度を向上することができる。

## 【 0 0 5 3 】

また、本発明は、英語に対しても適用が可能である。たとえば「1-800-1122」の読み方は、(1)「one eight hundred double one double two」、(2)「one eight oh oh eleven twelve」、(3)「one eight hundred one one two two」という各種の読み方がある。このような読み方についても同様に一会話内で維持されることが考えられる。このような 1 つの表記に対して複数の読みがある場合には、日本語に限られず、英語に対しても本発明を適用して 2 回目以降の認識精度を高めることができる。また、英語での文字についても同様である。たとえば「0」について、「zero」あるいは「oh」の複数の読みがある。このような場合も同様に本発明を適用できる。

## 【 0 0 5 4 】

また、英語では、数字の並びに対して、(1) 2 桁ずつ区切って桁読みする読み方、(2) 続けて数字を読む読み方(べた読み)、(3) 正式な桁読み、(4) 同じ数字の連続をdouble-xx triple-xx と表現する読み方等がある。たとえば「1999」について、「nineteen ninety nine」、「one nine nine nine」、「one thousand nine hundred ninety nine」、「one triple nine」、「nineteen double nine」、等の読み方がある。このような複数の読み方に対してはそれに対応した文法データが各々必要になる。1 回目の認識の際には全ての読み方を含む

文法を用意する必要があるが、2回目以降の認識では、本発明を適用して、不要と思われる文法データを削除できる。これにより、英語においても2回目以降の認識率を高めることができる。

## 【0055】

また、数字に限らず、通常の単語に対しても本発明を適用できる。たとえば、「青/ブルー」、「空飛ぶ円盤/ユーフオー/ユーエフオー」、「いりぐち/はいりぐち」など、同じ意味を持っていても複数の表現が可能な単語がある。このような単語に対しても本発明を提供して、2回目以降の認識では1回目で表現されなかった単語を削除して2回目以降の認識率を高めることができる。

## 【0056】

また、たとえば助動詞に対しても本発明を適用できる。たとえば「\*\*だ。」「\*\*です。」のように通常何れかの表現が安定して用いられる。このような場合にも、1回目で用いられなかった表現を削除して、2回目以降の認識率を向上できる。

## 【0057】

さらに、本発明は、動詞、形容詞その他の品詞に対しても複数の読み、表現がある場合にこれを適用することができ、認識精度を向上できる。

## 【0058】

## 【発明の効果】

本願で開示される発明のうち、代表的なものによって得られる効果は、以下の通りである。

## 【0059】

すなわち、音声認識における認識精度、特に複数の読みが与えられている文字、単語等の認識精度を向上できる。また、冗長性の少ない音たとえば日本語における数字や、異なる文字あるいは単語に類似の音を与えられている文字等の認識精度を向上できる。また、長母音付き一音節文字が短母音化されて繰り返される場合の認識精度を向上できる。

## 【図面の簡単な説明】

## 【図1】

本発明の一実施の形態である音声認識システムの概要を示したブロック図である。

【図 2】

実施の形態 1 で用いる音素表記表の一例を示す表図である。

【図 3】

実施の形態の音声認識方法の一例を示したフローチャートである。

【図 4】

減縮した文法データを示すリスト図面である。

【図 5】

さらに減縮した文法データを示すリスト図である。

【図 6】

実施の形態 2 の認識方法の一例を示したフローチャートである。

【図 7】

(a) は、実施の形態 3 の音声認識方法の一例を示したフローチャートであり、  
(b) は適用される文法データである。

【図 8】

非一様な確率分布を適用した文法データの一例を示すリスト図である。

【図 9】

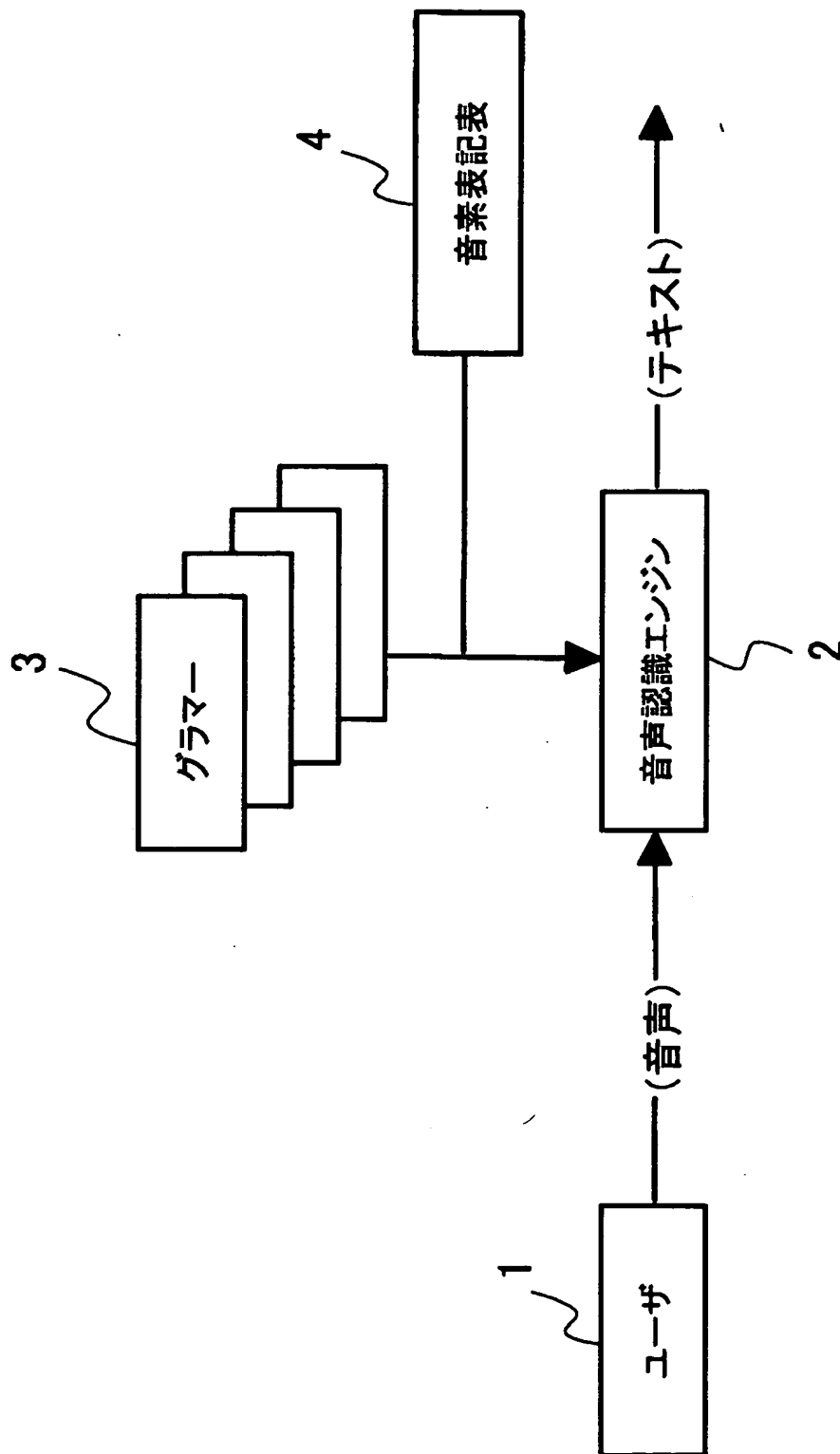
(a) は、4 桁の数字を認識する場合の文法を示すリスト図であり、(b) は、  
実用的な音素表記を示す表図であり、(c) は、(b) の音素表記を適用した  
文法データの例を示すリスト図である。

【符号の説明】

1 … ユーザ、 2 … 音声認識エンジン、 3 … 文法データ、 4 … 音素表記表。

【書類名】 図面

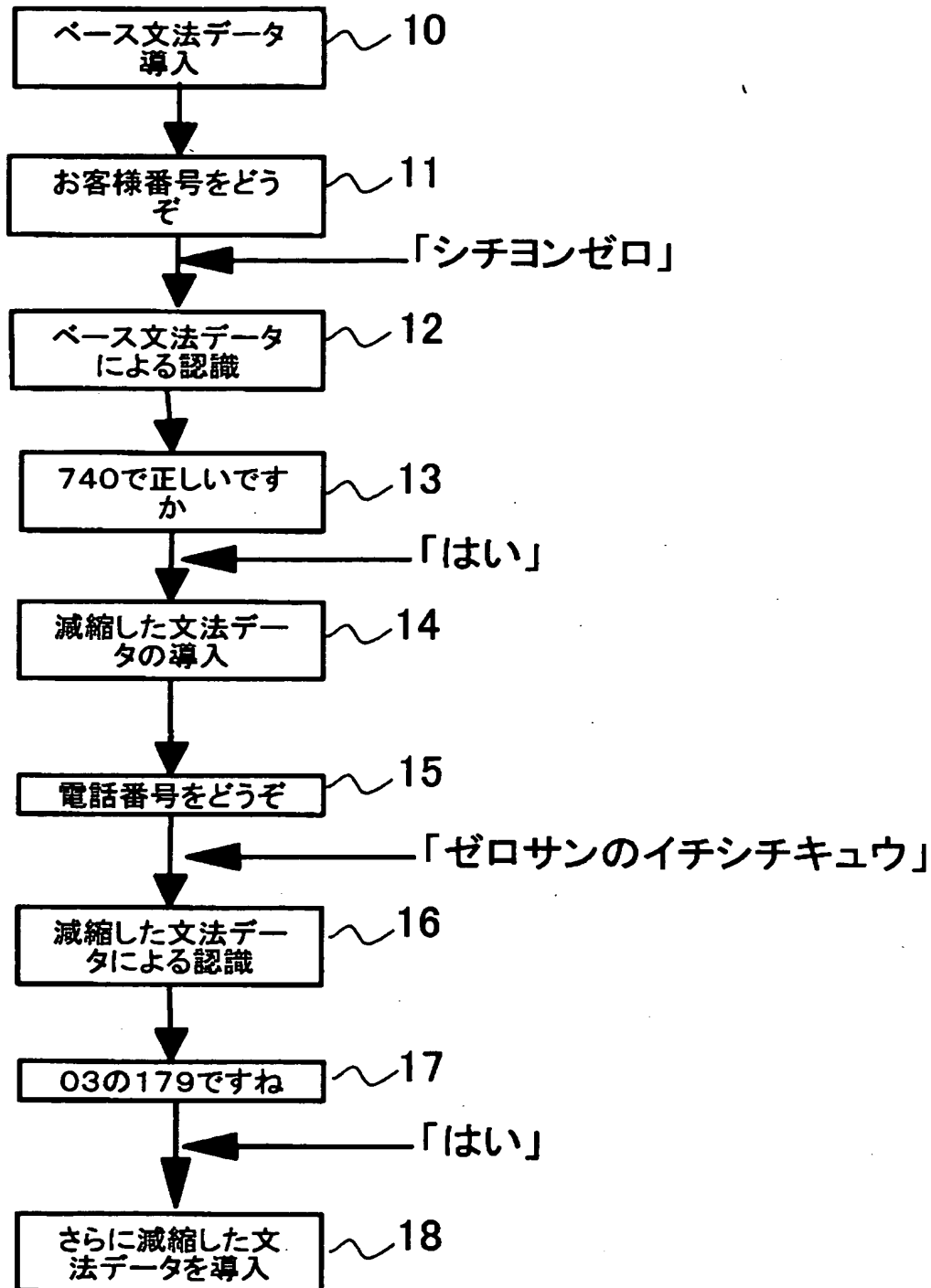
【図 1】



【図2】

0 (レイ)	ree
0 (レイ)	ree:
0 (レイ)	rei
0 (ゼロ)	zero
1	ichi
2	ni
2	nii
2	nii:
3	sa_n
4 (シ)	shi
4 (シ)	shii
4 (シ)	shii:
4 (ヨン)	yo_n
5	go
5	goo:
5	gou
6	roku
7 (シチ)	hichi
7 (ナナ)	nana
7 (シチ)	shichi
8	hachi
9 (ク)	ku
9 (キュウ)	kyuu
9 (キュウ)	kyuu:

【図 3】





【図 4】

<digits>::=<num1><num1><num1> .

<num1>=

0 |  
1 |  
2 |  
3 |  
4 |  
5 |  
6 |  
7 |  
8 |  
9 .

<0> = | z e r o .                      ← 22

<1> = | i c h i .

<2> = | n i | n i i | n i i .

<3> = | s a \_ n .

<4> = | y o \_ n .                      ← 21

<5> = | g o | g o o : | g o u .

<6> = | r o k u .

<7> = | h i c h i | s h i c h i .      ← 20

<8> = | h a c h i .

<9> = | k u | k y u u | k y u u : .

【図 5】

$\langle \text{digits} \rangle ::= \langle \text{num1} \rangle \langle \text{num1} \rangle \langle \text{num1} \rangle .$

$\langle \text{num1} \rangle =$

0 |

1 |

2 |

3 |

4 |

5 |

6 |

7 |

8 |

9 .

$\langle 0 \rangle = | \text{z e r o} .$

$\langle 1 \rangle = | \text{i c h i} .$

$\langle 2 \rangle = | \text{n i} | \text{n i i} | \text{n i i} .$

$\langle 3 \rangle = | \text{s a}_n .$

$\langle 4 \rangle = | \text{y o}_n .$

$\langle 5 \rangle = | \text{g o} | \text{g o o} : | \text{g o u} .$

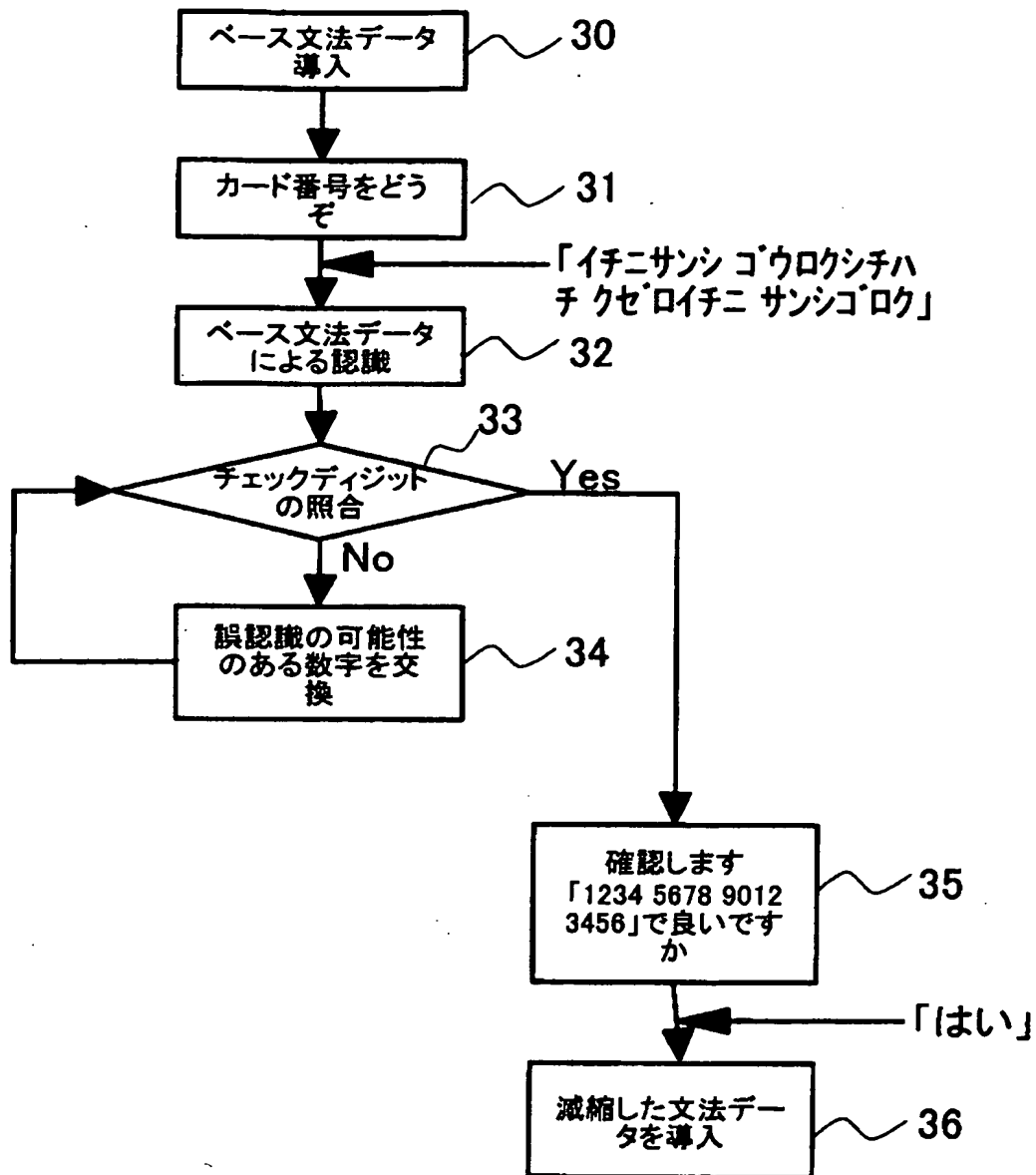
$\langle 6 \rangle = | \text{r o k u} .$

$\langle 7 \rangle = | \text{h i c h i} | \text{s h i c h i} .$

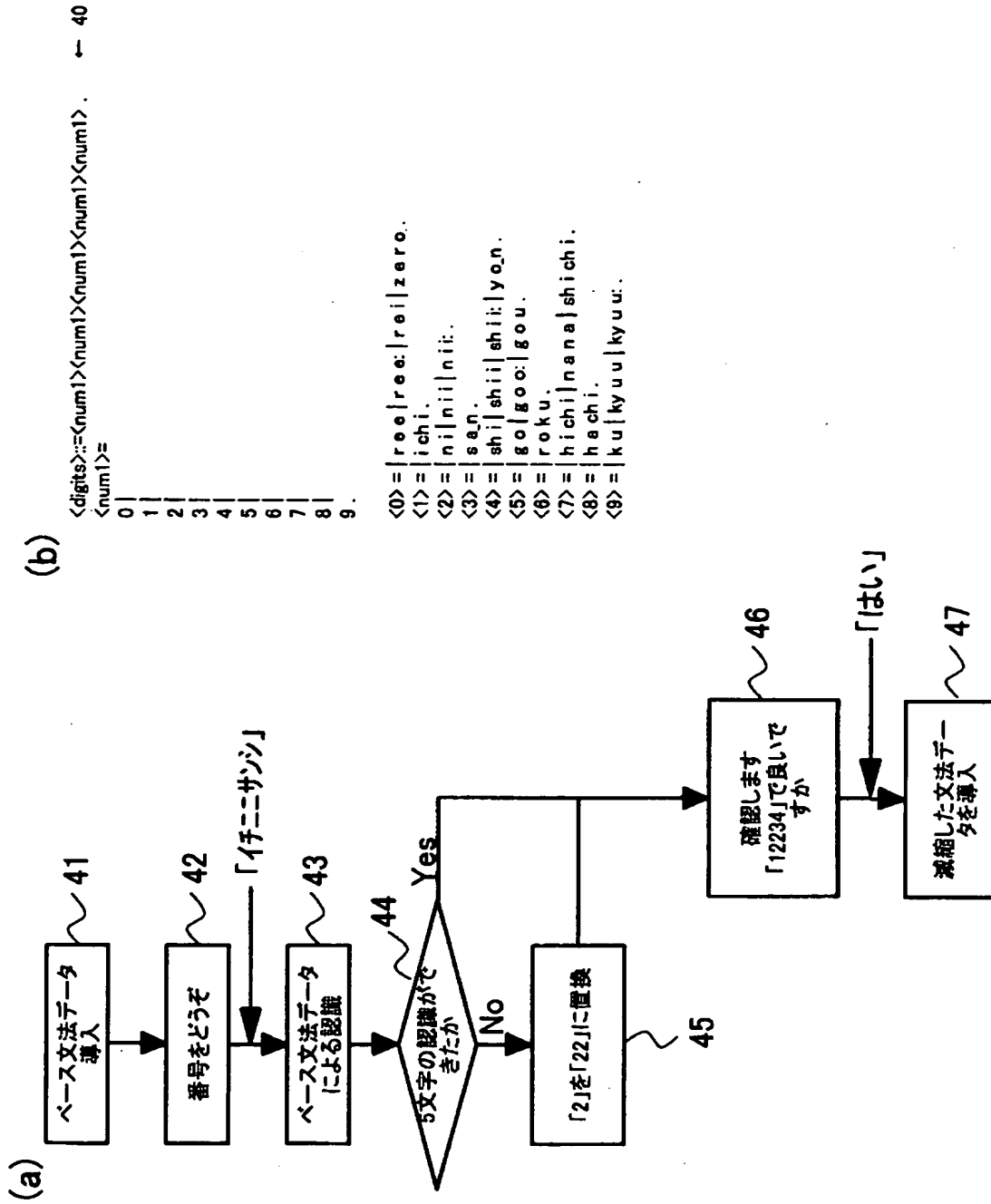
$\langle 8 \rangle = | \text{h a c h i} .$

$\langle 9 \rangle = | \text{k y u u} | \text{k y u u} : . \quad \leftarrow 23$

【図 6】



【図 7】



【図 8】

$\langle \text{digits} \rangle ::= \langle \text{num1} \rangle \langle \text{num1} \rangle .$

$\langle \text{num1} \rangle =$

0 |

1 |

2 |

3 |

4 |

5 |

6 |

7 |

8 |

9 .

$\langle 0 \rangle = | \text{zero} | \text{zero} | \text{zero} | \langle \text{rei} \rangle .$

$\langle 1 \rangle = | \text{ichi} .$

$\langle 2 \rangle = | \text{ni} | \text{nii} | \text{nii} : .$

$\langle 3 \rangle = | \text{san} .$

$\langle 4 \rangle = | \text{shi} | \text{shii} | \text{shii} : | \text{yo}_n .$

$\langle 5 \rangle = | \text{go} | \text{goo} : | \text{gou} .$

$\langle 6 \rangle = | \text{roku} .$

$\langle 7 \rangle = | \text{hichi} | \text{nana} | \text{shichi} .$

$\langle 8 \rangle = | \text{hachi} .$

$\langle 9 \rangle = | \text{ku} | \text{kyuu} | \text{kyuu} : .$

$\langle \text{rei} \rangle = | \text{ree} | \text{ree} : | \text{rei} .$

【図 9】

(a)		(b)		(c)	
<digits>::=<num1><num1><num1><num1><num1>		<digits>::=<num1><num1><num1><num1><num1>		<digits>::=<num1><num1><num1><num1><num1>	
<num1>=		<num1>=		<num1>=	
0	ree	0	rei	0	ree ree rei zero.
1	rei	0	zero	1	1 ichi.
2	ichi	1	ni	2	2 ni nii nit.
3	ni	2	nii	3	3 san.
4	nit	2	nit	4	4 shi shii shiti yon.
5	san	3	san	5	5 go goo gou.
6	shi	4	shi	6	6 roku.
7	shii	4	shii	7	7 hichi nana shichi.
8	shiti	4	shiti	8	8 hachi.
9	yon	4	yon	9	9 ku kyuu kyuu.
	go	5	go		
	goo	5	goo		
	gou	5	gou		
	roku	6	roku		
	hichi	7	hichi		
	nana	7	nana		
	shichi	7	shichi		
	hachi	8	hachi		
	ku	9	ku		
	kyuu	9	kyuu		
	kyuu	9	kyuu		

【書類名】 要約書

【要約】

【課題】 音声認識における認識精度、特に複数の読みが与えられてる文字、単語等の認識精度を向上する。

【解決手段】 同一人物は同一会話内で一定して同じ読みを維持するという傾向、つまり、たとえば「7」を「しち」と発音した人物は、その会話内で一貫して「しち」と発音し続ける傾向がある。この傾向を活用して、会話における1回目の応答においてその人物が使わない読みに対応する認識確率を下げ、2回目以降の認識を実行する。また、システムが音声合成により認識結果を復唱する場合に、既に認識した話者の読み方に対応して復唱する。たとえば話者が「7」を「しち」と発音した場合には、システムからの復唱時に「しち」と発音する。

【選択図】 図3

認定・付加情報

特許出願の番号	特願 2000-133973
受付番号	50000560703
書類名	特許願
担当官	塩崎 博子 1606
作成日	平成12年 6月16日

<認定情報・付加情報>

【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【復代理人】

【識別番号】	100112520
【住所又は居所】	神奈川県大和市中央林間3丁目4番4号 サクライビル4階 間山・林合同技術特許事務所
【氏名又は名称】	林 茂則

【選任した代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

【選任した復代理人】

【識別番号】	100110607
【住所又は居所】	神奈川県大和市中央林間3丁目4番4号 サクライビル4階 間山・林合同技術特許事務所
【氏名又は名称】	間山 進也

【選任した復代理人】

【識別番号】	100098121
--------	-----------

次頁有



認定・付加情報（続き）

【住所又は居所】 神奈川県大和市中央林間3丁目4番4号 サクラ  
イビル4階 間山・林合同技術特許事務所  
【氏名又は名称】 間山 世津子

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 1990年10月24日  
[変更理由] 新規登録  
住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)  
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション
  
2. 変更年月日 2000年 5月16日  
[変更理由] 名称変更  
住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)  
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション